Convexité et combinatoire

Jérôme Malick

Article invité

Convexité et combinatoire

Jérôme Malick
CNRS, Laboratoire Jean Kunztmann (LJK), Grenoble
jerome.malick@inria.fr

De nombreux problèmes dans l'industrie ou les services (en Finance, par exemple) peuvent être modélisés comme des problèmes d'optimisation de nature combinatoire. La résolution numérique de ces problèmes nécessite souvent la mise en oeuvre de méthodes spécifiques. On peut néanmoins isoler quelques idées fondamentales à la source de nombreuses approches ou techniques en optimisation combinatoire. Deux méthodes essentielles sont : **couper** et **majorer**. L'analyse convexe s'avère être un cadre théorique adapté à ces deux paradigmes. Ainsi l'optimisation convexe a permis d'élaborer de nouvelles méthodes - certaines pratiques, d'autres théoriques - qui complètent les méthodes plus combinatoires.

Chaque section ci-dessous présente brièvement une situation de l'**optimisation combinatoire** où l'**optimisation convexe** joue un rôle essentiel. Nous commencons par quelques mots sur l'apport de l'analyse convexe en génération de coupes. Nous laissons ensuite la « séparation », pour insister fortement sur la « majoration » en argumentant que l'optimisation convexe, et en particulier la dualité lagrangienne, fournit des outils théoriques et pratiques essentiels pour majorer. Nous évoquons ainsi tour à tour : les problèmes décomposables, le problème de la gestion de la production électrique en France, l'optimisation semidéfinie positive et un problème standard sur les graphes.

Ce article est destiné à des lecteurs venant de tous les horizons de la recherche opérationnelle ; son objectif est de leur donner l'envie de s'intéresser à l'analyse convexe [8]. La contrainte de place et le choix de faire un article de synthèse pour un public large forcent à commettre des approximations et se contenter d'une bibliographie très incomplète. Les experts sont donc avertis : cet article ne présente qu'une vision personnelle, partielle et partiale de ces idées de l'optimisation combinatoire.

Analyse convexe et coupes

Les méthodes génériques de type « branch and bound » pour résoudre de manière exacte des problèmes d'optimisation combinatoire gagnent considérablement en efficacité lorsqu'elles intègrent des techniques de « coupes » pour réduire les domaines réalisables. Lors de son exposé plénier à la conférence ROADEF en 2007, G. Cornuéjols expliquait que l'utilisation des coupes faisait gagner un facteur 1000 de temps de calcul pour résoudre avec CPLEX les problèmes linéaires en nombres entiers.

Une « coupe » (appelée aussi plan coupant) est un hyperplan affine qui sépare un point $\bar{x} \in \mathbb{R}^n$ d'un ensemble $S \subset \mathbb{R}^n$ (ou de son enveloppe convexe ce qui revient au même). Ceci correspond à la donnée d'une direction $d \in \mathbb{R}^n$ et d'un niveau $\alpha \in \mathbb{R}$ tel que

$$d^\top \bar{x} > \alpha \quad \text{et} \quad d^\top x \leq \alpha, \text{ pour tout } x \in S.$$

(Remarquons au passage qu'apparaissent ici les « fonctions supports » importantes dans toute l'analyse convexe [8]). Une coupe (d, α) est définie à une constante multiplicative près ; pour manipuler convenablement ces objets, [1] introduit le « polaire inverse », ouvrant ainsi les portes à l'analyse convexe.

Il est vrai que l'analyse convexe n'est pas nécessaire pour l'utilisation pratique des coupes. En effet, les techniques mathématiquement élaborées ne sont pas toujours les plus efficaces : les impératifs pratiques exigent souvent des coupes qui soient faciles à générer, avec peu de coefficients non nuls, et en plus avec des coefficients entiers de préférence. Il n'empêche que l'analyse convexe fournit un bon cadre pour manipuler les concepts liés aux coupes. C'est vrai en particulier pour les **coupes « disjonctives »** lorsqu'on veut séparer un point de l'union de deux polyèdres : une analyse de cette question a été réalisée récemment dans [6], et elle a inspiré un algorithme [4] pour calculer des coupes qui identifient des faces du polyèdre-union tout en maximisant la distance à $\bar{x}$ (au sens de la norme euclidienne).

Ce point de vue de l'analyse convexe permet aussi de montrer que les coupes « lift-and-project » (cas particulier de coupes disjonctives) s'interprètent comme des coupes qui maximisent la distance à $\bar{x}$ au sens d'une certaine norme (polyédrale). Il est probable que ces analyses aboutissent à des avancées techniques utiles dans les logiciels pour l'optimisation en nombres entiers.

Dualité et majoration

La dualité en optimisation linéaire est bien connue : les deux problèmes (« primal » et « dual »)

$$\left\{\begin{array}{l} \max \quad p^\top x \\ \quad Ax = b,\ x \geq 0 \end{array}\right. \qquad \left\{\begin{array}{l} \min \quad b^\top y \\ \quad A^\top y \geq p \end{array}\right. \tag{1}$$

admettent la même valeur optimale, et les variables duales admettent de plus une interprétation « économique » dans l'espace primal. Non-linéaires et même non-convexes, les problèmes d'optimisation admettent en fait des propriétés similaires dans un cadre bien plus général. Une bonne introduction sur la démarche de la dualité dite « lagrangienne » (on dit aussi **relaxation lagrangienne**) se trouve dans un numéro précédent de ce bulletin [10]. Ici, nous ne faisons qu'esquisser cette démarche, dans un cadre abstrait pour en faire ressortir quelques idées, puis nous l'illustrerons dans les deux sections suivantes.

Considérons un problème d'optimisation quelconque et séparons les contraintes en deux groupes : les contraintes symbolisées par $c(x) = 0$ seront « dualisées » (ou « relaxées »), les autres notées $x \in \mathcal{X}$ seront « gardées en dur ». La contrepartie de (1) dans cette situation abstraite s'écrit alors

$$\left\{\begin{array}{l} \max \quad f(x) \\ \quad c(x) = 0,\ x \in \mathcal{X} \end{array}\right. \qquad \left\{\begin{array}{l} \min \quad \theta(y) \\ \quad y \in \mathbb{R}^m \end{array}\right. \tag{2}$$

où la fonction « duale » est définie par

$$\theta(y) \ := \ \sup_{x \in \mathcal{X}} \ f(x) - y^\top c(x). \tag{3}$$

Définie comme un suprenum, θ est naturellement une fonction convexe et en général non-différentiable. Elle peut aussi prendre la valeur $+\infty$; cela donne implicitement des contraintes sur la variable duale y. On peut vérifier qu'on retrouve bien (1) dans le cas où $f(x) = p^\top x$, $c(x) = Ax - b$ et $\mathcal{X} = \mathbb{R}^n_+$.

Avec ces définitions, on montre facilement que chaque valeur $\theta(y)$ de la fonction duale fournit une majoration de la valeur optimale primale val(P). Le problème dual s'interprète donc comme le calcul du meilleur de ces majorants. Nous avons ainsi, pour toute variable duale y et toute variable primale x réalisable ($c(x) = 0$ et $x \in \mathcal{X}$),

$$f(x) \leq \mathrm{val(P)} \leq \mathrm{val(D)} \leq \theta(y).$$

Sans propriétés supplémentaires, il y a potentiellement un « saut dual », ce qui signifie que la seconde inégalité est stricte en général. Ainsi, bien qu'il n'y ait plus la dualité parfaite du cas linéaire (1), la dualité lagrangienne offre un **moyen systématique pour générer des majorants** du problème primal (qui peut être un problème combinatoire) : il suffit de résoudre (3) ou mieux encore de résoudre le problème dual, qui est toujours un problème d'**optimisation convexe** (non-différentiable). Il existe des méthodes efficaces adaptées à ces problèmes : la « génération de colonnes » ou la méthode des « faisceaux » ; voir [3] pour la comparaison de ces approches pour des problèmes combinatoires.

Utiliser (3) pour s'attaquer au problème primal de (2) n'est indiqué que si une hypothèse pragmatique est vérifiée : le problème (3) doit être facile à résoudre ! De plus la formulation d'un problème comme (2) dépend du choix des contraintes qui seront dualisées. Faire le bon choix de dualisation est tout un art : on peut voir que plus on dualise de contraintes, plus le majorant sera mauvais, mais plus le problème (3) sera facile a priori. La qualité de la borne obtenue et son intérêt dépendent ainsi d'un équilibre parfois subtile entre **précision et temps de calcul**.

Terminons cette section sur les concepts généraux en explicitant des cas où la dualité est parfaite, comme dans le cas linéaire (c'est-à-dire quand les deux problèmes (2) ont la même valeur optimale). Le résultat clé est que : si le problème primal de (2) est convexe (f convexe, c affine et $\mathcal{X}$ convexe), alors il n'y a pas de saut de dualité, sauf cas pathologiques. Sans hypothèse de convexité, nous pouvons quand même garantir (voir [13]) l'absence de saut sous la contrainte « unilatérale »

$$\text{tout } x \in \mathcal{X} \text{ vérifie } c(x) \leq 0.$$

Ce résultat explique les propriétés théoriques du semi-lagrangien, introduit avec succès par [2] pour résoudre le problème « p-median », qui revient à ajouter dans le problème primal la contrainte redondante $c(x) \leq 0$ à $\mathcal{X}$ et à utiliser la dualité lagrangienne standard.

Décomposition et production

Les problèmes « décomposables » forment une classe importante de problèmes d'optimisation se prêtant bien à la dualité lagrangienne. Avec les notations de la section précédente, cela correspond au cas où $\mathcal{X}$ est un produit cartésien et les fonctions f et c sont additives :

$$\left\{\begin{array}{l} \max \quad f(x) = \sum_j f_j(x_j) \\ \quad \sum_j c_j(x_j) = 0, \quad x \in \mathcal{X} = \Pi_j \mathcal{X}_j. \end{array}\right. \tag{4}$$

Le problème primal est ainsi une juxtaposition de problèmes « locaux », couplés par la contrainte c. Dualiser cette contrainte résulte donc en un problème totalement décomposé

$$\theta(y) = \sum_j \max_{x_j \in \mathcal{X}_j} f_j(x_j) - y_j c_j(x_j). \qquad (5)$$

Résoudre (4) par dualité lagrangienne s'appelle aussi décomposition par les prix : chaque y_j peut être interprété comme un prix marginal attribué à la violation de la j-ème contrainte. L'hypothèse pragmatique de la section précédente est valide puisque la résolution des problèmes locaux j est bien moins coûteuse que la résolution du problème initial ; il s'agit ensuite de trouver le meilleur prix y par un algorithme d'optimisation convexe.

Le problème de la **gestion de la production électrique** en France par EDF s'écrit sous la forme (4) : schématiquement, chaque centrale de production connait ses coût de production f_j et ses contraintes propres $\mathcal{X}_j$, et les contraintes couplantes (linéaires) consistent à assurer que les différentes demandes en électricité sont bien satisfaites (voir les détails sur le modèle dans [5]). La dualité lagrangienne est donc naturellement l'angle d'attaque choisi pour résoudre le problème de gestion de production (précisément : la gestion à court-terme) ; une méthode des faisceaux est utilisée depuis une dizaine d'années pour résoudre le problème dual.

Cette importante application industrielle des outils de l'optimisation combinatoire et de l'optimisation convexe est le fruit de plus de 20 ans de collaboration intensive entre EDF et la recherche académique française, en particulier C. Lemaréchal à l'INRIA. Nous poursuivons actuellement le travail sur deux points :

- l'amélioration de l'heuristique de [5] pour construire des solutions primales à partir des informations duales,
- l'accélération de la convergence de la méthode des faisceaux en utilisant la structure particulière de la fonction duale [14].

Optimisation SDP et graphes

Considérons les problèmes « quadratiques » : dans (2), les fonctions f, c sont linéaires ou quadratiques, et $\mathcal{X}$ est l'ensemble des zéros de fonctions quadratiques (voir par exemple le problème (6) ci-dessous). Pour ces problèmes, l'application du schéma de la dualité lagrangienne (voir [11] pour une approche systématique) fait apparaître des problèmes d'optimisation convexe bien particuliers : les problèmes d'optimisation « semidéfinie » ou « SDP ». Un problème SDP est un problème linéaire en une variable notée X qui est une matrice symétrique avec toutes ses valeurs propres positives (contrainte convexe de semidéfinie positivité). Voir par exemple le problème (7) ci-dessous. L'apparition dans les années 90 de l'optimisation SDP ([7] en est un article de référence) a eu un fort impact en optimisation combinatoire, en particulier pour la théorie de l'approximation. Les majorants SDP s'avèrent en effet être précis et le saut dual associé souvent controlable. Il existe de plus des algorithmes polynomiaux pour résoudre ces problèmes SDP (algorithmes de points intérieurs). En pratique, la question de l'efficacite des solveurs SDP est primordiale en vue de la recherche de l'équilibre entre précision et temps de calcul. La question est déjà de savoir quelles sont les tailles des instances qui peuvent effectivement être résolues par les solveurs SDP disponibles.

Illustrons brièvement ces idées sur le problème du **stable max**, un problème classique en théorie des graphes. Soit un graphe G ; intéressons-nous à $\alpha(G)$ la taille du plus grand sous-graphe de G sans aucune arête (le plus grand stable). En introduisant e le vecteur formé uniquement de 1, ce problème s'écrit comme le problème quadratique

$$\begin{cases} \max \quad e^\top x \\ \quad x_i x_j = 0 \quad \text{si } (i,j) \text{ est une arête de } G \\ \quad x_i \in \{0,1\} \quad \text{pour chaque sommet } i \end{cases} \qquad (6)$$

qui est bien un problème linéaire avec des contraintes quadratiques (après reformulation des contraintes binaires comme ${x_i}^2 = x_i$).

L'application automatique de la démarche de dualité lagrangienne donne une première formulation SDP, qui peut être ensuite simplifiée (par des calculs non-triviaux) pour s'écrire

$$\begin{cases} \max \quad \text{trace}(Xee^\top) \\ \quad X_{ij} = 0 \quad \text{pour toute arête } (ij) \\ \quad \text{trace}(X) = 1, \quad X \text{ SDP.} \end{cases} \qquad (7)$$

On appelle $\vartheta(G)$ (« le nombre de Lovász » [12]) la valeur optimale de ce problème d'optimisation SDP (linéaire en la matrice symétrique X). Par construction, $\vartheta(G)$ est une borne pour $\alpha(G)$; par ailleurs, on voit facilement que $\alpha(G) \leq \chi(\bar{G})$, où $\chi(\bar{G})$ est le nombre chromatique du graphe complémentaire $\bar{G}$. En fait, le théorème « du sandwich » nous dit plus précisément que

$$\alpha(G) \leq \vartheta(G) \leq \chi(\bar{G}).$$

La borne $\vartheta(G)$ donne ainsi une indication sur ces deux valeurs qui s'avèrent être NP-difficiles à calculer et même à approximer.

Il se trouve que pour résoudre ces problèmes (6), les solveurs SDP s'essoufflent vite lorsque la taille de G devient grande : par exemple certains graphes de la bibliothèque DIMACS [9] résistent aux solveurs SDP (ceux par points intérieurs, mais aussi ceux à base de faisceaux [16]). Le nombre de sommets de ces graphes n'est pas trop grand en général (typiquement plus petit que 1000), mais ils ont une densité équilibrée (G et $\bar{G}$ ont tous deux beaucoup d'arêtes). Voici deux exemples parmi d'autres : le graphe `p-hat-1000-3` a 1000 sommets, 127754 arêtes et 371746 arêtes dans le complémentaire ; le graphe `brock-800-1` en a lui respectivement 800, 112095 et 207505.

Dans cette situation, un travail récent [15] exploite une idée classique de l'optimisation convexe (la régularisation proximale de la fonction duale, ou lagrangien augmenté), pour introduire **une famille d'algorithmes pour l'optimisation SDP**. Ces nouvelles méthodes s'avèrent particulièrement adaptées aux problèmes issus de la théorie des graphes, qui ont beaucoup de contraintes (jusqu'à 100000) et des matrices pas trop grandes (≤ 1000). Cette approche par régularisation a permis de calculer le nombre $\vartheta(G)$ de plusieurs graphes de la bibliothèque DIMACS pour la première fois.

Remerciements. Merci à Florent Cadoux, Vincent Jost, Claude Lemaréchal, Julien Moncel et Frédéric Roupin, qui m'ont fait part de leurs commentaires sur ce document.

Références

[1] E. Balas, *Disjunctive programming*, Annals Discrete Mathematics, 5, 1979

[2] C. Beltran, C. Tadonki and J.-Ph.Vial, *Solving the p-median problem with a semi-Lagrangian relaxation*, Computational Optimization and Applications, 35, 2006

[3] O. Briant et al., *Comparison of bundle and classical column generation*, Mathematical Programming, 113, 2008

[4] F. Cadoux, *Computing deep facet-defining disjunctive cuts for mixed-integer programming*, to appear in Mathematical Programming, 2009

[5] L. Dubost, R. Gonzalez and C. Lemaréchal, *A primal-proximal heuristic applied to the French Unit-commitment problem*, Mathematical Programming, 104, 2005

[6] G. Cornuéjols and C. Lemaréchal, *A convex-analysis perspective on disjunctive cuts*, Mathematical Programming, 106, 2006

[7] M. Goemans and D. Williamson, *Improved approximation algorithms for max-cut and satisfiability problems using semidefinite programming*, Journal of the ACM, 6, 1995

[8] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer Verlag, Heidelberg, 2001

[9] D. Johnson and M. Trick, *Cliques, coloring and satisfiability : Second DIMACS Implementation Challenge*, American Mathematical Society, 1996

[10] C. Lemaréchal, *Borner, convexifier, relaxer : l'omniprésence de Lagrange*, Bulletin de la ROADEF, 8, 2002

[11] C. Lemaréchal and F. Oustry, *SDP relaxations and Lagrangian duality with application to combinatorial optimization*, Rapport de Recherche 3710, INRIA, 1999

[12] L. Lovász, *Semidefinite programs and combinatorial optimization*, in : Recent Advances in Algorithms and Combinatorics, Springer, 2003

[13] J. Malick, *The spherical constraint in boolean quadratic programs*, Journal of Global Optimization, 39, 2007

[14] J. Malick and S. Miller, *Newton methods for nonsmooth convex minimization*, Mathematical Programming, 104, 2005

[15] J. Malick, J. Povh, F. Rendl and A. Wiegele, *Regularization methods for semidefinite programming*, to appear in SIAM Journal on Optimization, 2009

[16] C. Helmberg and F. Rendl, *A spectral bundle method for semidefinite programming*, SIAM Journal on Optimization, 10, 2000